Proceedings of the 52nd Hawaii International Conference on System Sciences | 2019

Can Machines Learn to Detect Fake News? A Survey Focused on Social Media

Fernando Cardoso Durier da Silva
UNIRIO
fernando.durier@uniriotec.br

Rafael Vieira da Costa Alves
UNIRIO
rafael.alves@uniriotec.br

Ana Cristina Bicharra Garcia
UNIRIO
cristina.bicharra@uniriotec.br

Abstract

Through a systematic literature review method, in this work we searched classical electronic libraries in order to find the most recent papers related to fake news detection on social medias. Our target is mapping the state of art of fake news detection, defining fake news and finding the most useful machine learning technique for doing so. We concluded that the most used method for automatic fake news detection is not just one classical machine learning technique, but instead a amalgamation of classic techniques coordinated by a neural network. We also identified a need for a domain ontology that would unify the different terminology and definitions of the fake news domain. This lack of consensual information may mislead opinions and conclusions.

1. Introduction

Different from the beginning of the internet, we produce more data and information than we are able to consume. Consequently, it is possible that some misinformation or rumours are generated and spread throughout the web, leading other users to believe and propagate them, in a chain of unintentional (or not) lies. Such misinformation can generate illusive thoughts and opinions, collective hysteria or other serious consequences. In order to avoid such things to happen, specially closed to political events such as elections, researchers have been studying the information flow and generation on social medias in the last years, focusing on subjects as opinion mining, users relationship, sentiment analysis, hatred spread, etc.

Based on a systematic review of recent literature published over the last 5 years, we synthesized different views dealing with fake news. We wanted to investigate machine learning applications to detect fake news, focusing on the characteristics of the different approaches and techniques, conceptual models for detecting fake news and the role of bots (cognitive agents) in this context as they have gained great popularity in the last three years.

In order to answer our questions and show the results of our work, we will present the definition of misinformation, hoax, fake news and its main common concept, meanwhile, systematically review a set of machine learning and natural language processing/nlp techniques used to detect such kind of information. We conclude outlining the challenges and research gaps in current state-of-art of automatic fake news detection.

2. Survey Methodology

We followed the systematic literature review for software engineering, SLR method, as prescribed in [1] and [2]. In this section we describe, step by step, the way we select and filter papers, analyze the research proposals and contributions in the papers, as well as, synthesized the results. Our search was guided by our research questions to understand the state of art of automatic detection of fake news.

For automating the SLR process, we used the Parsifal tool, an online collaborative SLR tool that allowed us to define a set of keywords, key research questions, query string, inclusion and exclusion criteria, and define the set of search sources.

We defined that our work should cover the aspects of automatic detection of fake news, therefore we chose the as our population the online newspapers, as our comparison we chose machine learning techniques, our outcome is fake news detection and our context is of an academic survey, also we defined the following keywords and synonyms on Parsifal [1], as we can see on Table 1

Parsifal is already integrated to IEEE, ACM and ScienceDirect digital libraries sources. This feature facilitates the selection phase of the literature review. Our choices were limited to what Parsifal's SLR automatic tool offered. However, the results offered from those, were great in terms of quality and quantity.

[1] https://parsif.al/

URI: https://hdl.handle.net/10125/59713
ISBN: 978-0-9981331-2-6

HICSS

Table 1. Keywords used on search.

Keywords	Synonyms	Related To
Detection	Stance, Tracking, Veracity	Intervention
Fake News	Automated Fact Checking, Disinformation, Hoax, Misbehaviour, Misinformation, Rumor	Outcome
Machine Learning	Artificial Intelligence, ML, Natural Language Processing, NLP	Comparison
Social Media	Facebook, News, Newspaper, Twitter	Population

A tradeoff offered by the automation, that in the end, didn't show to be a problem for us.

The query generated by our chosen keywords was ”(”Detection” OR ”Stance” OR ”Tracking” OR ”Veracity”) AND (”machine learning” OR ”Artificial Intelligence” OR ”ML” OR ”Natural Language Processing” OR ”NLP”) AND (”Fake news” OR ”Automated Fact checking” OR ”disinformation” OR ”Hoax” OR ”misbehaviour” OR ”misinformation” OR ”Rumor”)”, retrieving a total of 1093 articles.

Our first selection criteria considered publication year. Fake news is a recent topic of interest, so we just consider papers published in the last 5 years. Older papers were only selected whenever they were important for understanding definitions as we were looking for the most recent and most up to date techniques the authors are using. Secondly, we preferred the experimental papers with actual data and results which used any machine learning, artificial intelligence or automated decision making algorithm. Finally, we preferred papers related to politics and deeply read the more robust (having deep description of techniques, experiments and concepts) ones.

On the study selection step of our research, we limited the set of papers to be only the ones published from 2008 to 2018. Then, we discarded all the papers without machine learning/nlp approaches or not about fake news detection, resulting in the remainder of 169 papers.

Having selected our study set, we analyzed those papers by first reading the abstract, introduction, theoretical references and conclusions in order to separate the most interesting ones. Having those most interesting ones, we proceed to a second deeper reading over those, in order to review their techniques, definitions, theoretical background, type of study and results.

The important aspects we were looking for in our readings can be described as follows: Definition of Fake News, Scientific Work Type, Natural Language Processing Techniques Used, Machine Learning Techniques Used, Which Step of the Detection Pipeline the Work Focused On and Which Social Aspects were Taken in Consideration for the Research.

The definition of fake news used by each author is important to us, as the term became very diffused by politicians, journalists, researchers and users throughout the medias, even more after the USA 2016 Election Events. The fake news were sometimes identified as rumours, hoaxes, spams, misinformation or simply fake news. Although, all of those keywords have exclusive attributes that separate them in their respective meaning group, e.g. a hoax being a more satirical misinformation that has no intention to manipulate opinions, but, to make a fun critic, or e.g. rumour being a unverified fact that is easily spread, easily adopted and used for opinion manipulation either for good or for malicious purposes, they all, converge to the same sense and concept, that is what we concluded on section 3. We needed to find a common definition to all in order to, not only for conceptual enlightenment, but for assertiveness in our revision, and meta-modeling reference of future works, as this would be the foundation for experiments in the area.[3][4][5]

The scientific work type is important to separate in niches what have been done as research in this so recent area. And if any survey were found, how we can improve the state-of-art knowledge in our work.

Both the Natural Language Process and Machine Learning Techniques, were our main focus in this works, as we are interested in which are the main used methodologies and techniques that can automatically detect/classify a micro-blog, social media or newspaper entry as a fake news or not, which features would be needed to be included in the classifiers and models, which were not relevant and so on.

We paid attention also to which steps of the detection pipeline the authors chose to intensify their works on, in order to know what is the most difficult part of classifying those entry in categories such as fake or not fake. And how they overcame their challenges on doing so, in order to in future works we would be able to overcome such challenges also.

During our preliminaries reading of theme-related papers, we found that many authors used social aspects of the entries in order to better classify them. Those aspects varied from comments, entry sharing, relationships between consumers of those entries to the writers profile information and pictures. We classified those aspects as relevant also, hoping to find new and better practices on how to use those external contextual information in favor of better predictions or better modeling.

3. Theoretical Reference

There are many definitions of fake news on the literature. In addition, the media has overused the term ”fake news” in many different contexts and with distinct intents, which aggravates the problem of understanding what characterizes a given story as fake news. In this section we extend the definitions from [6] to characterize the properties of fake news. Then, we present some papers definition of fake news and its relations with some related concepts, that are commonly wrongly used as interchangeable by the media.

3.1. Publisher

We define the **Publisher** as the entity that provides the story to a public. For example, the publisher can be an user in a micro-blogging service like Twitter, a journalist in an online newspaper, or an organization in its own website. Note, that the publisher may or may not be the author of the story.

In the case that the publisher is the author of the story, Kumar and Shah[6] classifies the author based on its **intent** into **misinformation**, if the author has not the intent to *deceive*, or into **disinformation**, if the author has the intent to *deceive*.

When the publisher is just spreading the story, ie. republishing content from other story, then we can classify them in bots and normal users.

3.2. Content

We define **Content** as the main information provided by the publisher in the story. At the moment of publication, the veracity of this information can be true, false or unknown. If the veracity is unknown, then it can be classified as a **rumor**, according with the definition of rumor from Zubiaga et al[7] as *”an item of circulating information whose veracity status is yet to be verified at the time of posting.”*

The information can also be classified as factual, opinion or mixed. Opinion based information have no ground truth, in contrast with factual information, where the facts can be verified against a ground truth. In the case of factual, usually the content is a **claim**, made by the publisher. The veracity of the **claim** is the object of study of *automated fact-checking*, which has a recent report from Nieminen et al [8]

3.3. Extra media

In addition to the content, the story may include some medias like picture, video, audio. The use of medias unrelated to the content, with the objective of increasing the will of the reader to access the content, is considered clickbaiting and is researched on the works of cite cite cite. Also, the author may modify the content of the media to give emphasis to a point of view.

3.4. Fake News Definition and its Impact on Society

The authors use different names to define the same concept that can be observed in our works reviewed. They call it misinformation, rumour, hoax, malicious trend, spam or fake news, but all converge to the same semantic meaning, that is of an information that is unverified, of easy spread throughout the net, with the intention of either block the knowledge construction (by spreading irrelevant or wrong information due to lack of knowledge of the theme) or either manipulate the readers opinion. [5][9][10][11][12][13][14]

The majority of works consider it to be consequence of excessive marketing strategies or political manipulation. It should be observed though, that some authors consider the chance of those stream and spread of misinformation being unintentional sometimes, and happening due to cultural shock (e.g., Nepal Earthquake case described on [11]) or unconscious acts.

In this work we will utilize the definition of fake news from Shu [15], which is ”a news article that is intentionally and verifiable false”. Note that this definition shares similarities our definition of a publisher with the intent to deceive and false factual content. However this definition is simplistic, since it does not cover half truths, opinion based contents, and humorous stories, like satires.

Due to the popularization of artificial intelligence and related areas of cognitive computing, the number of bots has exploded throughout the network. In this section, we will explore their role in the rumours and misinformation spreading.[16][17]

Some authors argue that the creation of bots, the cognitive agents, would be more harmful to the information recovery process, due to the fact that they would intensify the propagation of misinformation, hoaxes and spams. [17]

However, we can see in [18] that this is more or less a truthful statement. As they discovered through their experiments that in fact, bots would increase the misinformation propagation indeed, but, they also would increase the true information propagation as well. Concluding that bots, are not misinformation spreaders but, just information spreaders not favoring one type of it, but accelerating propagation of any kind of information.

4. Social Medias

Through our readings we found that most of the works use the social medias and micro-blogs as their main source of analysis. This is due to the increasing use of social networks by everyone, like Facebook, Twitter and Google+. Mainly Twitter and Sina Weibo. [19][20][21]

In addition, the microblogging platforms usually provide an API (Application Programming Interface) to query and consume its data. The APIs usually provides the content of the platform in structured data or plain text, thus reducing the preprocessing step that is commonly used with web crawlers used to filter the information of interest from web pages. [22][16][23]

Another reason for this is that most of newspapers are just too serious and express more a generic political opinion compared to the social networks that express individual opinions of many different users with different beliefs, contexts and cultural backgrounds. Also it is very difficult to find an expressive newspaper that diffuse rumours and fake news, as the assurance of information quality is part of the newspaper's main process.

Nowadays, the politician context is being heavily influenced by the fake news dissemination and existence.

To the point of some countries being lawfully prepared for such scenario, in Brazil for example, minister Luiz Fux, in a seminary said that if a Brazilian election has been biased by fake news it would be annulled. [24]

Some social medias in some countries are beginning to think on new strategies of combating this, by contracting third-party enterprises to help in defining/tagging which information is fake news or not, e.g. For Facebook the strategy was to use checker agencies that monitor news and classify them as fake or not fake, specifically the agency A Lupa uses the following scale: (1) True; (2) True, however, needing more explanation; (3) Too recent to affirm anything; (4) Exaggerated; (5) Contradictory; (6) Unbearable; e (7) False. When this interviews was done, the Facebook team affirmed that this strategy lessened in 80% the Fake News ”organically” generated in the US by use of similar agencies there.[25]

In the other hand WhatsApp in Brazil limited the number of messages with the same content that can be shared by the same user, is using a Artificial Intelligence to detect abuses and harass messages and like Facebook using third-party agencies to check and classify news. Also, the WhatsApp team trained and showed the capabilities of their app to the current president candidates and their communication team in an attempt to avoid possible use of the app for fake news spread. [26]

5. Machine Learning

There are numerous classifiers on the literature. From simple Random Forests and Naive Bayes to SVMs. In this section we will present the different kind of models, preprocessing techniques and datasets used on the literature.

5.1. Public Datasets and Challenges

For reference and future works, we selected a few public datasets and challenges in the area.

In the year of 2017, two challenges were proposed by the community, namely the RumorEval (SemEval 2017 Shared Task 8) and the Fake News Challenge. The former had two subtasks, one for stance detection of a reply to a news, and another for classifying the news as true or false. The latter is just a stance detection of a news, which classifies the reply of a news in agrees, disagrees, discussing and unrelated.

There are numerous sites for manual fact checking on the web. Two of the most popular are snopes.com and factcheck.org. In addition, there are specialized sites, for specialized domains like politics, like politifact.com. In contrast, there are also numerous of sites, like theonion.com, that publish news explicitly declared fake. Many of these sites are publishing these news as a satire, humorous, or as a critic. Many papers generated their dataset from these two sources. The fact checking as ground truth to true news and satire online journals as ground truth to false news.

Wang provided the LIAR dataset[27], composed of statements made by public figures, annotated with its veracity, extracted from the site polifact.com For research focused on rumors, there is the PHEME dataset, by Zubiaga et al.[28]. This dataset groups a number of tweets in rumor threads, and associate them with news events.

5.2. Preprocessing

Most of papers use preprocessing steps in order to either increase the tax of correct rate or to have a faster processing[29][10][30].

There are works that focus on automatically detect the starting point of the rumours' stream, by topologic exploration. The authors of [31], proposed an algorithm to do so and obtained good results (compared to the other ones they tested against) finding the origin of the rumour news, furthermore, they discovered key features that tended to appear on those kind of tweets and use them in future works to pre-clusterize the scrapped tweets and agilize the origin tracking process and fasten misinformation classification.

5.3. NLP Features

Many papers used sentiment analysis to classify the polarity of a news[32][33][34][35][10]. Some used sentiment lexicons, which demand a lot of human effort to build and maintain, and built a supervised learning based classifier. Some papers which use such approach of sentiment analysis as feature for final classifiers, use chain models like Hidden Markov Models or Artificial Neural Network in order to infer sentiment.

The usage of other techniques based on syntax is relatively low. Papers mainly use parsing, pos-tagging and named entity types. On the other hand, the use of semantics are more common. Many papers used lexicons as external knowledge about words, creating lists of words based on properties of interest. For example, swear words, subjective words, and sentiment lexicons. Commonly used lexicons are WordNet and Linguist Inquiry and Word Count (LIWC)

Another use of semantics on fake news detection is the use of language modelling. Some papers used n-grams as baselines for comparisons with their handcrafted features. Others used n-grams as features to their classifiers[36]. More recent papers[37][32] used word embeddings for language modelling, mainly the ones that are constructing a classifier using unsupervised learning. Word embeddings is a family of language models, where a vocabulary is mapped to a high dimension vector. These language models assign a real-valued vector to each word in the vocabulary, with the objective that words close in meaning will be close in the vector space. One of most commonly used model is the word2vec from Mikolov et al. [38], which uses a neural network to estimate the vectors.

5.4. Social and Content Features

We grouped the features found in the classifiers in sets based on the source of the feature. The first set groups features based on social media attributes (#likes, #retweets, #friends). The second set has features based on the content of the news (punctuations, word embeddings, sentiment polarity of words).

As we could see in [39], there is a preference for more classical classification algorithms that heavily focus on the linguistic aspects. But also, we can see the increased usage of new methods that aggregate different, yet on the same context, features to give better results and insights, such as Network Topology Analysis Models and Artificial Neural Networks that explore the link between users and other meta information provided by the social media predefined data structure.

Some authors propose to classify the social medias entries as fakes by analysing its interaction between users. Based on this, we found interesting the proposed work [40]. Motivated by the collaborative aspect of nowadays web2.0, and by the of swarm intelligence (or collective intelligence), the authors explore how is given the process of forming a collective knowledge from interactions of social networks users, in an event they name as social swarm.

Using a german dataset from an online gamer community, they apply statistics and linguistic analysis to extract text data to pass it through a set of classical machine learning algorithms for classification, those being Nive Bayes, K-Nearest Neighbours (KNN), Decision Tree and Support Vector Machine (SVM). To counterpoint this classical analysis, they try an approach of what they define as ant algorithm.

The Ant Algorithm works much like an ant colony. The news are sprayed with pheromones, while there is such in the vicinity of the data acquired, the algorithm operates until the pheromone evaporate, increasingly predicting and updating its error ratio, till the thread of total pheromones is totally evaporated. A much interesting and ludicrous approach of such problem. The algorithm only classify the news as Positive or Negative, however for their purposes, it is just what they wanted.

Compared to other classical methods, heuristics and algorithms, this one showed to be the best one with the lesser error rate of all. In our scenario, it could be applied to detect fake news, hoax, rumours or misinformation by modifying its classification function, as most of works that handle fake news detection depends on interaction analysis, and this new algorithm proved to be much more efficient to this task than its classical counterparts, even though its implementation would be more complex.

5.5. Models

Different from what we were expecting the authors in the literature didn't used simple and classical machine learning models, like Naive Bayes, Decision Tree, Support Vector Machine, etc. Instead they tried a combination of those in a more powerful and, as far as their results shown, more accurate composite model.

In order to achieve such composition the authors recurred to a model which is gained much popularity on the last years, the Artificial Neural Networks (ANN).

6. Open Challenges and Future Work

Multimodal classifiers: Most of news embed medias (videos, pictures) in the content, but it may not be related to the content and is there only for marketing purposes (clickbaiting). There is a work that focuses on classifying tweets by analyzing its memes that could be helpful in such an effort ([41]), in fact they also focus on an pre-tagging step based on recurrent terms that goes altogether whenever posting such memes.

Another open challenge we found was that there is still uncertainty over the real intents of a tweet. Due to linguistic resources like metaphors, euphemisms and sarcasm the real intention of that microblog entry would be very well understandable for a human reader, however, the machine is not necessarily trained during its KDD process to differ language forms, but to tag or classify what is written, or by cross-checking with a predefined dictionary of pre-classified terms. So, we find that an interesting research gap to attack in future works would be the disambiguation of tweets intents.

7. Conclusion

Although many researchers argue that the social media and such information obtained from its metrics, is a key-feature for election prediction, others argue that this approach is too simplistic due to the lack of certainty over the real goal of political discussion on such social medias, as many tend to be satirical and not really serious, or the lack of an algorithmic and logic formalism preliminary definitions and even arguing that the good performance/scoring of the election winners on social networks per say would not be enough to establish a causality relationship to the urn victory. [39] Also, there is a work [32] which creates an attention based ANN with textual, social and image information sources and applied it on twitter and Weibo datasets, achieving 75% accuracy.

On the social information propagation used as preprocessing step, we come to conclude that it is a very favorable approach, since it helps on identifying key-features to be used as enrichment on classifying process, helps on finding the starting point of spread and pretag it as a rumour spreader (which proved to decrease the propagation rate from that point forward) and helps on mapping the external contextual elements from the microblogging entries.

As we reviewed, the preferred methods of handling the problem of fake news, rumours, misinformation detection is the machine learning approach, mainly, involving composite classifiers that are in fact neural networks composed by classical classification algorithms that heavily focus on lexical analysis of the entries as main features for prediction, and the usage of external contextual information (e.g. topologic distribution of microblogging entries, users profiles, social media metrics, etc.) to improve classification results as a preliminary process step of such models.

The natural language processing approaches are used on the literature more as a preliminary step than a solution per say. We are not saying that it is not relevant, we are arguing that it is more a part of the final machine learning solutions than what we expected.

About the usage of bots, we can conclude that they can be viewed as catalysts of information propagation, either for good purposes or bad ones. They don't favor a type of entry, but instead help propagating it faster due to its computational capabilities that surpass those of a human being, and due to its popularity that turned them to be easier to manufacture and easier to use and being adopted by users. Of course, there are many ways to improve their information validation characteristics in future works, but, it would demand a lot of preprocessing of those external contextual elements we saw on topologic analysis of entries.

Different from many surveys we read[42][7], we came to conclude that the current state of art of automatic detection of fake news is of using composite network analysis approaches on the machine learning techniques choices, we came to conclude that a new more generic concept of fake news could be defined so it would ease future metamodelling of the entry object and enable better generalistic misinformation detecting agents to be manufactured.

References

[1] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of systems and software*, vol. 80, no. 4, pp. 571–583, 2007.

[2] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Information and software technology*, vol. 55, no. 12, pp. 2049–2075, 2013.

[3] P. T. Metaxas, E. Mustafaraj, and D. Gayo-Avello, “How (Not) to Predict Elections,” pp. 165–171, IEEE, Oct. 2011.

[4] R. Ushigome, T. Matsuda, M. Sonoda, and J. Chao, “Examination of Classifying Hoaxes over SNS Using Bayesian Network,” pp. 606–608, IEEE, Nov. 2017.

[5] E. C. Tandoc Jr, Z. W. Lim, and R. Ling, “Defining fake news a typology of scholarly definitions,” *Digital Journalism*, pp. 1–17, 2017.

[6] S. Kumar and N. Shah, “False information on web and social media: A survey,” *CoRR*, vol. abs/1804.08559, 2018.

[7] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, “Detection and Resolution of Rumours in Social Media: A Survey,” *ACM Computing Surveys*, vol. 51, pp. 1–36, Feb. 2018.

[8] S. Nieminen and L. Rapeli, “Fighting misperceptions and doubting journalists objectivity: A review of fact-checking literature,” *Political Studies Review*, p. 1478929918786852, 2018.

[9] L. Zheng and C. W. Tan, “A probabilistic characterization of the rumor graph boundary in rumor source detection,” pp. 765–769, IEEE, July 2015.

[10] J. A. Ceron-Guzman and E. Leon-Guzman, “A Sentiment Analysis System of Spanish Tweets and Its Application in Colombia 2014 Presidential Election,” pp. 250–257, IEEE, Oct. 2016.

[11] J. Radianti, S. R. Hiltz, and L. Labaka, “An Overview of Public Concerns During the Recovery Period after a Major Earthquake: Nepal Twitter Analysis,” pp. 136–145, IEEE, Jan. 2016.

[12] S. Ahmed, R. Monzur, and R. Palit, “Development of a Rumor and Spam Reporting and Removal Tool for Social Media,” pp. 157–163, IEEE, Dec. 2016.

[13] M. Rajdev and K. Le, “Fake and Spam Messages: Detecting Misinformation During Natural Disasters on Social Media,” pp. 17–20, IEEE, Dec. 2015.

[14] I. Y. R. Pratiwi, R. A. Asmara, and F. Rahutomo, “Study of hoax news detection using nave bayes classifier in Indonesian language,” pp. 73–78, IEEE, Oct. 2017.

[15] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” *SIGKDD Explorations*, vol. 19, no. 1, pp. 22–36, 2017.

[16] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, and F. Menczer, “The DARPA Twitter Bot Challenge,” *Computer*, vol. 49, pp. 38–46, June 2016.

[17] C. Shao, G. L. Ciampaglia, O. Varol, A. Flammini, and F. Menczer, “The spread of fake news by social bots,” *arXiv preprint arXiv:1707.07592*, 2017.

[18] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[19] Q. Wang, T. Lu, X. Ding, and N. Gu, “Think twice before reposting it: Promoting accountable behavior on Sina Weibo,” pp. 463–468, IEEE, May 2014.

[20] K. Wu, S. Yang, and K. Q. Zhu, “False rumors detection on Sina Weibo by propagation structures,” pp. 651–662, IEEE, Apr. 2015.

[21] G. Liang, J. Yang, and C. Xu, “Automatic rumors identification on Sina Weibo,” pp. 1523–1531, IEEE, Aug. 2016.

[22] C. Buntain and J. Golbeck, “Automatically Identifying Fake News in Popular Twitter Threads,” pp. 208–215, IEEE, Nov. 2017.

[23] M. D. Conover, B. Goncalves, J. Ratkiewicz, A. Flammini, and F. Menczer, “Predicting the Political Alignment of Twitter Users,” pp. 192–199, IEEE, Oct. 2011.

[24] P. Felipe, “Justia eleitoral desafiada por fake news,” *Agłncia Brasil*.

[25] J. Valente, “Redes sociais adotam medidas para combater fake news nas eleies,” *Agłncia Brasil*.

[26] B. Capelas, “Whatsapp anuncia planos para tentar combater ’fake news’ no brasil,” *Estado*.

[27] W. Y. Wang, “” liar, liar pants on fire”: A new benchmark dataset for fake news detection,” *arXiv preprint arXiv:1705.00648*, 2017.

[28] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, and P. Tolmie, “Analysing how people orient to and spread rumours in social media by looking at conversational threads,” *PloS one*, vol. 11, no. 3, p. e0150989, 2016.

[29] A. Boutet, D. Frey, R. Guerraoui, A. Jegou, and A.-M. Kermarrec, “WHATSUP: A Decentralized Instant News Recommender,” pp. 741–752, IEEE, May 2013.

[30] S. Fong, S. Deb, I.-W. Chan, and P. Vijayakumar, “An event driven neural network system for evaluating public moods from online users’ comments,” pp. 239–243, IEEE, Feb. 2014.

[31] Sahana V P, A. R. Pias, R. Shastri, and S. Mandloi, “Automatic detection of rumoured tweets and finding its origin,” pp. 607–612, IEEE, Dec. 2015.

[32] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, “Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs,” pp. 795–816, ACM Press, 2017.

[33] N. Hassan, F. Arslan, C. Li, and M. Tremayne, “Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster,” pp. 1803–1812, ACM Press, 2017.

[34] S. Vosoughi, M. . Mohsenvand, and D. Roy, “Rumor Gauge: Predicting the Veracity of Rumors on Twitter,” *ACM Transactions on Knowledge Discovery from Data*, vol. 11, pp. 1–36, July 2017.

[35] J. Ross and K. Thirunarayan, “Features for Ranking Tweets Based on Credibility and Newsworthiness,” pp. 18–25, IEEE, Oct. 2016.

[36] P. Dewan, S. Bagroy, and P. Kumaraguru, “Hiding in plain sight: Characterizing and detecting malicious Facebook pages,” pp. 193–196, IEEE, Aug. 2016.

[37] A. P. B. Veyseh, J. Ebrahimi, D. Dou, and D. Lowd, “A Temporal Attentional Model for Rumor Stance Classification,” pp. 2335–2338, ACM Press, 2017.

[38] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Advances in neural information processing systems*, pp. 3111–3119, 2013.

[39] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," in *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, ASIST '15, (Silver Springs, MD, USA), pp. 82:1–82:4, American Society for Information Science, 2015.

[40] C. Kaiser, A. Piazza, J. Krockel, and F. Bodendorf, "Are Humans Like Ants? – Analyzing Collective Opinion Formation in Online Discussions," pp. 266–273, IEEE, Oct. 2011.

[41] E. Ferrara, M. JafariAsbagh, O. Varol, V. Qazvinian, F. Menczer, and A. Flammini, "Clustering memes in social media," pp. 548–555, ACM Press, 2013.

[42] S. Kumar, M. Jiang, T. Jung, R. J. Luo, and J. Leskovec, "MIS2: Misinformation and Misbehavior Mining on the Web," pp. 799–800, ACM Press, 2018.